

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2001, 18:24:52 ; Search time: 113.17 Seconds
(without alignments)
225.225 Million cell updates/sec

Title: us-09-373-230-5

Sequence: 1 TTYGARGARGATGAVCC 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_ov1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pi1:*
13: gb_pi2:*
14: gb_pi3:*
15: gb_pi4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pi1:*
48: em_rod:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v12:*
59: gb_v12:*
60: gb_v12:*
61: gb_v12:*
62: gb_v12:*
63: gb_v12:*
64: gb_v12:*
65: gb_v12:*
66: gb_v12:*
67: gb_v12:*
68: gb_v12:*
69: gb_v12:*
70: gb_v12:*
71: gb_v12:*
72: gb_v12:*
73: gb_v12:*
74: gb_v12:*
75: gb_v12:*
76: gb_v12:*
77: gb_v12:*
78: gb_v12:*
79: gb_v12:*
80: gb_v12:*
81: gb_v12:*
82: gb_v12:*
83: gb_v12:*
84: gb_v12:*
85: gb_v12:*
86: gb_v12:*
87: gb_v12:*
88: gb_v12:*
89: gb_v12:*
90: gb_v12:*
91: gb_v12:*
92: gb_v12:*
93: gb_v12:*
94: gb_v12:*
95: gb_v12:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|---------------------|
| 1 | 15.4 | 90.6 | 17 | 9 | AR072047 Sequence |
| 2 | 15.4 | 90.6 | 471 | 9 | AR072044 Sequence |
| 3 | 15.4 | 90.6 | 471 | 10 | E13264 Mouse cDNA |
| 4 | 15.4 | 90.6 | 471 | 10 | E13264 Mouse cDNA |
| 5 | 15.4 | 90.6 | 471 | 10 | E14257 Murine mRNA |
| 6 | 15.4 | 90.6 | 471 | 10 | E14760 Murine mRNA |
| 7 | 15.4 | 90.6 | 471 | 45 | E17139 Murine mRNA |
| 8 | 15.4 | 90.6 | 471 | 45 | E10609 Mouse cDNA |
| 9 | 15.4 | 90.6 | 471 | 45 | E11744 CDNA encodi |
| 10 | 15.4 | 90.6 | 471 | 56 | E12010 CDNA encodi |
| 11 | 15.4 | 90.6 | 572 | 94 | MM066244 Artificial |
| | | | | | U66244 Mus musculu |

| | | | | | | |
|----|------|------|--------|----|-----------|---------------------|
| 12 | 15.4 | 90.6 | 866 | 94 | MUSIGIEPP | DA9949 Mouse mRNA |
| C | 15.4 | 90.6 | 35143 | 90 | AP000229 | AC000229 Homo sapi |
| 13 | 15.4 | 90.6 | 59955 | 77 | AC080029 | AC080029 Homo sapi |
| C | 15.4 | 90.6 | 95155 | 88 | AC078950 | AC078950 Homo sapi |
| C | 15.4 | 90.6 | 100000 | 90 | AP000144 | AP000144 Homo sapi |
| C | 17 | 90.6 | 100634 | 90 | AP001594 | AP001594 Homo sapi |
| C | 18 | 90.6 | 123631 | 91 | HS22F01 | AL109967 Homo sapi |
| C | 19 | 90.6 | 202103 | 87 | AC010176 | AC010176 Homo sapi |
| C | 20 | 90.6 | 340000 | 90 | AP001695 | AP001695 Homo sapi |
| C | 21 | 90.6 | 355 | 54 | G01707 | G01707 chicken STS |
| 22 | 14.4 | 84.7 | 710 | 94 | MMHBEF14 | U93192 Mus musculus |
| C | 23 | 14.4 | 84.7 | 93 | CNS01GZU | AL143803 Anopheles |
| C | 24 | 14.4 | 84.7 | 97 | CNS07CNO | AL143404 T7 end of |
| C | 25 | 14.4 | 1084 | 53 | CNS07OT9 | AL424051 clone AZ0 |
| C | 26 | 14.4 | 1188 | 5 | AF318610 | AF318610 Caenorhab |
| C | 27 | 14.4 | 1718 | 12 | AF046922 | AF046922 Colletotr |
| C | 28 | 14.4 | 1971 | 4 | AB035057 | AB035057 Drosophila |
| C | 29 | 14.4 | 2094 | 4 | AB035062 | AB035062 Drosophila |
| C | 30 | 14.4 | 2178 | 94 | MUSHEGF04 | AB035069 Drosophila |
| C | 31 | 14.4 | 2237 | 4 | AB035069 | AB035069 Drosophila |
| C | 32 | 14.4 | 2268 | 4 | AB035062 | AB035062 Drosophila |
| C | 33 | 14.4 | 2282 | 4 | AB035061 | AB035061 Drosophila |
| C | 34 | 14.4 | 2305 | 4 | AB035065 | AB035065 Drosophila |
| C | 35 | 14.4 | 2919 | 12 | AF024633 | AF024633 Magnapor |
| C | 36 | 14.4 | 3172 | 12 | AF015753 | AF015753 Magnapor |
| C | 37 | 14.4 | 3522 | 4 | AB035058 | AB035058 Drosophila |
| C | 38 | 14.4 | 4438 | 8 | AF085184 | AF085184 Gallus ga |
| C | 39 | 14.4 | 84.7 | 13 | RICPPA | D32136 Rice endoge |
| C | 40 | 14.4 | 84.7 | 5 | CEK10C8 | AF155773 Gibberell |
| C | 41 | 14.4 | 21077 | 5 | CEK10C8 | 274474 Caenorhabd |
| C | 42 | 14.4 | 21644 | 92 | HS2A1KLT | 281330 Human DNA s |
| C | 43 | 14.4 | 22092 | 4 | CELR0637 | AC006637 Caenorhab |
| C | 44 | 14.4 | 24361 | 5 | CELR02C2 | AF030307 Caenorhab |
| C | 45 | 14.4 | 27748 | 5 | CELR34C03 | AF100662 Caenorhab |

ALIGNMENTS

| | | | | | | |
|------------|--|------------|-----|-----|-------------|--|
| RESULT | 1 | | | | | |
| LOCUS | AR072047 | 17 bp | DNA | PAT | 18-FEB-2000 | |
| DEFINITION | Sequence 5 from patent US 5912324. | | | | | |
| ACCESSION | AR072047 | | | | | |
| VERSION | AR072047.1 | GI:7222935 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | Unknown. | | | | | |
| ORGANISM | Unknown. | | | | | |
| REFERENCE | 1 (bases 1 to 17) | | | | | |
| AUTHORS | Okamura,H., Tanimoto,T., Torigoe,K., Kunikata,T., Taniguchi,M., Kohno,K. and Kurimoto,M. | | | | | |
| TITLE | Interferon-gamma (IFN-gamma.) inducing factor (IGIF, IL-18) purified from murine liver | | | | | |
| JOURNAL | Patent: US 5912324-A 5 15-JUN-1999; | | | | | |
| FEATURES | Location/Qualifiers | | | | | |
| source | 1..17 | | | | | |
| BASE COUNT | 4 a | 2 c | 4 g | 3 t | 4 others | |
| ORIGIN | /organism="unknown" | | | | | |

Query Match 90.6%; Score 15.4; DB 9; Length 17;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGARATGAYCC 17
 |||||
 Db 1 TTYGARGARATGAYCC 17

RESULT 2
 AR072044

| | | | | | |
|------------|--|------------|------|-------|-------------|
| LOCUS | AR072044 | 471 bp | DNA | PAT | 18-FEB-2000 |
| DEFINITION | Sequence 1 from patent US 5912324. | | | | |
| ACCESSION | AR072044 | | | | |
| VERSION | AR072044.1 | GI:7222932 | | | |
| KEYWORDS | | | | | |
| SOURCE | Unknown. | | | | |
| ORGANISM | Unknown. | | | | |
| REFERENCE | 1 (bases 1 to 471) | | | | |
| AUTHORS | Okamura,H., Tanimoto,T., Torigoe,K., Kunikata,T., Taniguchi,M., Kohno,K. and Kurimoto,M. | | | | |
| TITLE | Interferon-gamma (IFN-gamma.) inducing factor (IGIF, IL-18) purified from murine liver | | | | |
| JOURNAL | Patent: US 5912324-A 1 15-JUN-1999; | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..471 | | | | |
| BASE COUNT | 162 a | 91 c | 92 g | 125 t | 1 others |
| ORIGIN | /organism="unknown" | | | | |

Query Match 90.6%; Score 15.4; DB 9; Length 471;
 Best Local Similarity 76.5%; Pred. No. 4.5e+02;
 Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGARATGAYCC 17
 |||||
 Db 244 TTYGARGARATGAYCC 260

| | | | | | |
|------------------|--|------------|------|-------|-------------|
| RESULT | 3 | | | | |
| LOCUS | E13264 | 471 bp | DNA | PAT | 24-JUN-1998 |
| DEFINITION | Mouse cDNA encoding a protein that induces to produce | | | | |
| ACCESSION | E13264 | | | | |
| VERSION | E13264.1 | GI:3252069 | | | |
| KEYWORDS | JP 1997157180-A/2. | | | | |
| SOURCE | Mus sp. | | | | |
| ORGANISM | Mus sp. | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurgnathia; Muridae; Murinae; Mus. | | | | |
| AUTHORS | Torigoe,K., Tanimoto,T., Fukuda,S. and Kurimoto,M. | | | | |
| TITLE | AGENT FOR SENSITIVE DISEASE | | | | |
| JOURNAL | Patent: JP 1997157180-A 2 17-JUN-1997; | | | | |
| COMMENT | HAYASHIBARA BIOCHEM LAB INC | | | | |
| OS | Mus sp. (mouse) | | | | |
| PN | JP 1997157180-A/2 | | | | |
| PD | 17-JUN-1997 | | | | |
| PE | 24-JAN-1996 JP 1996028722 | | | | |
| PF | 10-MAR-1995 JP 95P 78357, 29-SEP-1995 JP 95P 274988, PR | | | | |
| PI | 04-OCT-1995 JP 95P 279906 | | | | |
| PT | TORIGOE KAKUTU, TANIMOTO TADAQ, FUKUDA SHIGETSU, PI | | | | |
| KUJIMOTO MASASHI | | | | | |
| PC | A61K38/00, A61K38/00, A61K38/00, A61K38/00, C07K14/52, | | | | |
| PC | C07K14/54, | | | | |
| PC | C07K14/55; | | | | |
| CC | strandedness: Double; | | | | |
| CC | topology: Linear; | | | | |
| CC | Feature is identified by similarity; | | | | |
| FH | Key | | | | |
| FT | Location/Qualifiers | | | | |
| FT | source | | | | |
| FT | 1..471 | | | | |
| FT | /organism="Mus sp." | | | | |
| FT | /tissue="liver" | | | | |
| FT | /product="interferon-gamma inducer protein" | | | | |
| FT | Location/Qualifiers | | | | |
| FT | 1..471 | | | | |
| FT | /organism="Mus sp." | | | | |
| FT | /db_xref="taxon:10095" | | | | |
| BASE COUNT | 162 a | 91 c | 92 g | 125 t | 1 others |

| FH | Key | Location/Qualifiers |
|----|--------|---------------------|
| FH | | |
| FT | source | 1. .471 |

| | | | | |
|--------------------------|--------|--------------------|-----------|-------------|
| Query Match | 90.6%; | Score 15.4; | DB 45; | Length 471; |
| Best Local Similarity | 76.5%; | Pred. No. 4.5e+02; | | |
| Matches 13; Conservative | 4; | Mismatches 0; | Indels 0; | Gaps 0 |

```

OY      1  TTYGARGARATGAGACC 17
      ||:||||:||||:|
DB      244  TTTGAGGAATGATGCC 260

RESULT  9
E12010  standard; RNA; ROD: 471 BP.
XX      E12010;
XX      E12010;
XX      E12010.1
XX      07-OCT-1997 (Rel. 52, Created)
XX      02-SEP-2000 (Rel. 65, Last updated, Version 2)
DE      cDNA encoding mouse polypeptide which introduce interferon-gamma product in
DE      immunocompetent cell.
XX      JP 1996231598-A/2.
XX      Mus musculus (house mouse)
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX      [1]
XX      1-471
RA      Kunikata T., Taniguchi M., Kono K., Kurimoto M.;
RT      "MONOCLONAL ANTIBODY";
RT      Patent number JP1996231598-A/2, 10-SEP-1996.
RL      HAYASHIBARA BIOCHEM LAB INC.
XX      XX
XX      OS      Mus musculus (mouse)
XX      PN      JP 1996231598-A/2
XX      CC      PD      10-SEP-1996
XX      CC      PF      23-FEB-1995 JP 1995058240
XX      CC      PI      KUNIKATA TOSHIO, TANIGUCHI MUTSUOKO, KONO KEIZO,
XX      CC      PI      KURIMOTO MASASHI
XX      CC      PC      C07K16/24, C07K1/16, C07K1/22, C07K1/26, C07K1/30,
XX      CC      PC      C07K1/34, C12N5/10
XX      CC      PC      C12N15/02, C12P21/08, G01N33/53, G01N33/577//A61K38/21, A61K39/395,
XX      CC      PC      C12P21/08,
XX      CC      PC      C12R1:91);
XX      CC      CC      strandedness: Double;
XX      CC      CC      topology: Linear;
XX      CC      CC      hypothetical: No;
XX      CC      CC      anti-sense: No;
XX      CC      FH      Key      Location/Qualifiers
XX      CC      CC      FT      source      1..471
XX      CC      CC      FT      /organism="Mus musculus"
XX      CC      CC      FT      /tissue_type="liver"
XX      CC      CC      FT      /tissue_type="liver"
XX      CC      FH      Key      Location/Qualifiers
XX      CC      FH      FT      source      1..471
XX      CC      FH      FT      /db_xref="taxon:10090"
XX      CC      FH      FT      /organism="Mus musculus"
XX      SO      Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other:

Query Match      90.6%; Score 15.4; DB 45; Length 471;
Best Local Similarity 76.5%; Pred. No. 4.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY      1  TTYGARGARATGAGACC 17
      ||:||||:||||:|
DB      244  TTTGAGGAATGATGCC 260

RESULT  10

```

```

ASIGIF LOCUS 471 bp DNA SYN 08-NOV-1996
DEFINITION Artificial sequence DNA for cytokine interferon-gamma inducing
factor (IGIF).
ACCESSION Y09278.1 GI:1666283
VERSION Y09278
KEYWORDS cytokine; interferon-gamma inducing factor (IGIF).
SOURCE synthetic construct.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 471)
AUTHORS Schluesener, H.J.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 471)
AUTHORS Schluesener, H.J.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1996) H.J. Schluesener, Inst.f. Hirnforschung,
Calwer Str. 3, D-77076 Tuebingen, FRG
FEATURES
SOURCE 1..471
/organism="synthetic construct"
/db_xref="taxon:32630"
<1..>471
/codon_start=1
/transl_table=1
/product="mouse cytokine interferon-gamma inducing factor"
/protein_id="CAA70473.1"
/db_xref="GI:1666284"
/translation="MPCRLHCTTAVIRNINDQVLFVDRKROPVEDMTIDDSASPPQ
RLIYKWDSEVRCIAYTLISYDSDKMSPLSKNKLISFEEMDPEDIIDISDLIFQ
KVPGRHNMEESSLYGHFLACQKEDDAFLILKKDEGDKSVNFTLTNHS"
BASE COUNT 148 a 100 c 130 g 93 t
ORIGIN

Query Match      90.6%; Score 15.4; DB 56; Length 471;
Best Local Similarity 76.5%; Pred. No. 4.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY      1  TTYGARGARATGAGACC 17
      ||:||||:||||:|
DB      244  TTTGAGGAATGAGACC 260

RESULT  11
MMU66244 572 bp mRNA ROD 18-MAR-1997
LOCUS Mus musculus interferon-gamma inducing factor mRNA, partial cds.
DEFINITION MMU66244
ACCESSION U06244
VERSION U06244.1 GI:1561735
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 572)
AUTHORS Rothe, H., Jenkins, N.A., Copeland, N.G. and Kolb, H.
TITLE Active stage of autoimmune diabetes is associated with the
expression of a novel cytokine, IGIF, which is located near Idd2
JOURNAL J. Clin. Invest. 99 (3), 469-474 (1997)
MEDLINE 97174346
REFERENCE 2 (bases 1 to 572)
AUTHORS Rothe, H., Copeland, N.G. and Kolb, H.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-1996) Diabetes Research Institute, Auf'm
Hennekamp 65, Dueseldorf 40225, Germany
FEATURES
SOURCE 1..572
/organism="Mus musculus"
/strain="non obese diabetic (NOD)"
/db_xref="taxon:10090"
/chromosome="9"
/tissue_type="pancreas"

```

CDS

1. >572
/note="IGIF"
/codon_start=1
/product="interferon gamma inducing factor"
/protein_id="BAB49753.1"
/db_xref="GI:1561736"
/translation="MAAASEDCVNFKEKMFIDNTLYFIPENGDLSDNFGRLHCTT
AVIRINDQVLFVDKRPVEDMTDIDQASAPQRLITYKDSYVGLAVTLVSKD
SKMSTLSCNKNKISFEEDMDPENIDIDIOSLDFQKRVPGHNKMEFESSLYEGHFLAC
OKEDDAFKLLKKKDENGDKSVFTLTNLHQS"

BASE COUNT 196 a 111 c 113 g 152 t

ORIGIN

Query Match 90.6%; Score 15.4; DB 94; Length 572;
Best Local Similarity 76.5%; Pred. No. 4.6e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYCC 17
||:|:|:|:|:|:|:|:|:|
Db 349 TTTGAGGAATGATGCC 365

RESULT 12
MUSIGIFPP 866 bp mRNA ROD 10-FEB-1999
LOCUS Mouse mRNA for IGIF precursor polypeptide, complete cds.
DEFINITION D49949
VERSION D49949.1 GI:1064822
KEYWORDS IGIF precursor polypeptide.
SOURCE Mus musculus liver cDNA to mRNA, clone pMUGF37B-5.
ORGANISM Mus musculus

REFERENCE
AUTHORS
JOURNAL
1 (bases 1 to 866)
Okamura, H.
Direct Submission
Submitted (29-MAR-1995) to the DDBJ/EMBL/GenBank databases. Haruki
Okamura, Hyogo College of Medicine, Department of Bacteriology;
Mukogawa 1-1, Nishinomiya, Hyogo 663, Japan (tel:0798-45-6111)

REFERENCE
AUTHORS
JOURNAL
2 (bases 1 to 866)
Okamura, H., Tsutsumi, H., Komatsu, T., Yutsudo, M., Hakura, A.,
Tanihara, T., Toriige, K., Okura, T., Nukada, Y., Hattori, K., Akita, K.,
Namba, M., Tanabe, F., Konishi, K., Fukuda, S. and Kurimoto, M.
Cloning of a new cytokine that induces IFN-gamma production by T
cells

TITLE
JOURNAL
MEDLINE
NATURE 378 (6552), 88-91 (1995)
FEATURES
SOURCE

Location/Qualifiers
1. 866
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="liver"
1. 866
/note="Interferon-gamma inducing factor (IGIF) mRNA"
165. 743
/codon_start=1
/product="IGIF precursor polypeptide"
/protein_id="BAB08705.1"
/db_xref="GI:1064823"
/translation="MAAASEDCVNFKEKMFIDNTLYFIPENGDLSDNFGRLHCTT
AVIRINDQVLFVDKRPVEDMTDIDQASAPQRLITYKDSYVGLAVTLVSKD
SKMSTLSCNKNKISFEEDMDPENIDIDIOSLDFQKRVPGHNKMEFESSLYEGHFLAC
OKEDDAFKLLKKKDENGDKSVFTLTNLHQS"

CDS
165. 743
/note="Interferon-gamma inducing factor (IGIF) mRNA"

polya_site 262 a 187 c 187 g 230 t
BASE COUNT
ORIGIN

Query Match 90.6%; Score 15.4; DB 94; Length 866;
Best Local Similarity 76.5%; Pred. No. 4.7e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYCC 17
||:|:|:|:|:|:|:|:|:|
Db 513 TTTGAGGAATGATGCC 529

RESULT 13
AP000229/c 35143 bp DNA PRI 17-MAR-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q21.1-q21.2, clone: T1539,
L156-APP region, complete sequence.
DEFINITION
AP000229
ACCESSION AP000229.2 GI:7262565
VERSION
KEYWORDS
SOURCE HTG.
Homo sapiens DNA, clone: T1539.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
JOURNAL
1 (bases 1 to 35143)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y. and Sakaki, Y.
Homo sapiens 35,143 genomic DNA of 21q21.1-q21.2
Published only in Database (1999) In press
2 (bases 1 to 35143)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y. and Sakaki, Y.
Direct Submission
Submitted (13-MAY-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamiharu, Kanagawa 228-8555, Japan
(E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
Tel: 81-42-778-9923, Fax: 81-42-778-9924)
On Mar 17, 2000 this sequence version replaced gi:4835598.
Sequence updated (15-Mar-2000)
The sequence is a part of the data (ACCESSION No. AP000136 -
AP000145).

COMMENT
The sequencing project is supported by Japan Science Technology
Corporation (JST) and The Institute of Physical and Chemical
Research (RIKEN).

FEATURES
SOURCE
Location/Qualifiers
1. 35143
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="T1539"
/map="21q21.1-q21.2"

BASE COUNT 10208 a 7036 c 7565 g 10334 t
ORIGIN

Query Match 90.6%; Score 15.4; DB 90; Length 35143;
Best Local Similarity 76.5%; Pred. No. 6.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYCC 17
||:|:|:|:|:|:|:|:|:|
Db 25608 TTTGAGGAATGATGCC 25592

RESULT 14
AC080029 59955 bp DNA HTG 25-OCT-2000
LOCUS Homo sapiens chromosome 4 clone CTD-2353K2 map 4, LOW-PASS SEQUENCE
DEFINITION
AC080029
ACCESSION AC080029.1 GI:10280784
VERSION
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 59955)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 4, clone CTD-235K2

2 (bases 1 to 59955)

Birren, B., Linton, J., Nussbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Barna, N., Bastien, V., Bede, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lacombe, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, C., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneses, L., Mihova, T., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Turrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project information

Center project name: L10743

Center clone name: 2353_K_2

NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 645: contig of 645 bp in length
* 646 745: gap of 100 bp
* 746 1464: contig of 719 bp in length
* 1465 1564: gap of 100 bp
* 1565 2315: contig of 751 bp in length
* 2316 2415: gap of 100 bp
* 2416 3059: contig of 644 bp in length
* 3060 3159: gap of 100 bp
* 3160 3877: contig of 718 bp in length
* 3878 3977: gap of 100 bp
* 3978 4722: contig of 745 bp in length
* 4723 4822: gap of 100 bp
* 4823 5579: contig of 757 bp in length
* 5580 5679: gap of 100 bp
* 5680 6315: contig of 636 bp in length
* 6316 6415: gap of 100 bp
* 6416 7281: contig of 665 bp in length
* 7281 7380: gap of 100 bp
* 7381 8134: contig of 754 bp in length
* 8135 8234: gap of 100 bp
* 8235 8980: contig of 746 bp in length
* 8981 9080: gap of 100 bp
* 9081 9763: contig of 683 bp in length
* 9764 9863: gap of 100 bp

9864 10479: contig of 616 bp in length
* 10480 10579: gap of 100 bp
* 10580 11203: contig of 624 bp in length
* 11204 11303: gap of 100 bp
* 11304 12043: contig of 740 bp in length
* 12044 12143: gap of 100 bp
* 12144 12846: contig of 703 bp in length
* 12847 12946: gap of 100 bp
* 12947 13678: contig of 732 bp in length
* 13679 13778: gap of 100 bp
* 13779 14542: contig of 764 bp in length
* 14543 14642: gap of 100 bp
* 14643 15366: contig of 724 bp in length
* 15367 15466: gap of 100 bp
* 15467 16160: contig of 694 bp in length
* 16161 16260: gap of 100 bp
* 16261 17065: contig of 805 bp in length
* 17066 17165: gap of 100 bp
* 17166 17914: contig of 749 bp in length
* 17915 18014: gap of 100 bp
* 18015 18715: contig of 701 bp in length
* 18716 18815: gap of 100 bp
* 18816 19436: contig of 621 bp in length
* 19437 19536: gap of 100 bp
* 19537 20176: contig of 640 bp in length
* 20177 20276: gap of 100 bp
* 20277 21013: contig of 737 bp in length
* 21014 21113: gap of 100 bp
* 21114 21872: contig of 759 bp in length
* 21873 21972: gap of 100 bp
* 21973 22634: contig of 662 bp in length
* 22635 22734: gap of 100 bp
* 22735 23340: contig of 606 bp in length
* 23341 23440: gap of 100 bp
* 23441 24159: contig of 719 bp in length
* 24160 24259: gap of 100 bp
* 24260 24993: contig of 734 bp in length
* 24994 25093: gap of 100 bp
* 25094 25806: contig of 713 bp in length
* 25807 25906: gap of 100 bp
* 25907 26519: contig of 613 bp in length
* 26520 26619: gap of 100 bp
* 26620 27253: contig of 634 bp in length
* 27254 27353: gap of 100 bp
* 27354 28058: contig of 705 bp in length
* 28059 28158: gap of 100 bp
* 28159 28834: contig of 676 bp in length
* 28835 28934: gap of 100 bp
* 28935 29594: contig of 660 bp in length
* 29595 29694: gap of 100 bp
* 29695 30458: contig of 764 bp in length
* 30459 30558: gap of 100 bp
* 30559 31323: contig of 765 bp in length
* 31324 31423: gap of 100 bp
* 31424 32170: contig of 747 bp in length
* 32171 32270: gap of 100 bp
* 32271 33037: contig of 767 bp in length
* 33038 33137: gap of 100 bp
* 33138 33845: contig of 708 bp in length
* 33846 33945: gap of 100 bp
* 33946 34637: contig of 692 bp in length
* 34638 34737: gap of 100 bp
* 34738 35427: contig of 690 bp in length
* 35428 35527: gap of 100 bp
* 35528 36173: contig of 646 bp in length
* 36174 36273: gap of 100 bp
* 36274 36921: contig of 648 bp in length
* 36922 37021: gap of 100 bp
* 37022 37736: contig of 715 bp in length
* 37737 37836: gap of 100 bp
* 37837 38624: contig of 788 bp in length
* 38625 38724: gap of 100 bp
* 38725 39480: contig of 756 bp in length

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2001, 10:13:29 ; Search time 1105.85 seconds
(without alignments)
226.716 Million cell updates/sec

Title: US-09-373-230-5
Perfect score: 17
Sequence: 1 TTYGARGARATGAYCC 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pi1:*
13: gb_pi2:*
14: gb_pi3:*
15: gb_pi4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_rod:*
33: em_htg_other:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pi1:*
48: em_pi2:*
49: em_pi3:*
50: em_pi4:*
51: em_pi5:*
52: em_pi6:*
53: em_pi7:*
54: em_pi8:*
55: em_pi9:*
56: em_pi10:*
57: em_pi11:*
58: em_pi12:*
59: em_pi13:*
60: em_pi14:*
61: em_pi15:*
62: em_pi16:*
63: em_pi17:*
64: em_pi18:*
65: em_pi19:*
66: em_pi20:*
67: em_pi21:*
68: em_pi22:*
69: em_pi23:*
70: em_pi24:*
71: em_pi25:*
72: em_pi26:*
73: em_pi27:*
74: em_pi28:*
75: em_pi29:*
76: em_pi30:*
77: em_pi31:*
78: em_pi32:*
79: em_pi33:*
80: em_pi34:*
81: em_pi35:*
82: em_pi36:*
83: em_pi37:*
84: em_pi38:*
85: em_pi39:*
86: em_pi40:*
87: em_pi41:*
88: em_pi42:*
89: em_pi43:*
90: em_pi44:*
91: em_pi45:*
92: em_pi46:*
93: em_pi47:*
94: em_pi48:*
95: em_pi49:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 15.4 | 90.6 | 17 | AR072047 | AR072047 Sequence |
| 2 | 15.4 | 90.6 | 471 | AR072044 | AR072044 Sequence |
| 3 | 15.4 | 90.6 | 471 | E13264 | E13264 Mouse cDNA |
| 4 | 15.4 | 90.6 | 471 | E14257 | E14257 cDNA encodi |
| 5 | 15.4 | 90.6 | 471 | E14760 | E14760 cDNA encodi |
| 6 | 15.4 | 90.6 | 471 | E17139 | E17139 Murine mR |
| 7 | 15.4 | 90.6 | 471 | E10609 | E10609 Mouse cDNA |
| 8 | 15.4 | 90.6 | 471 | E11744 | E11744 cDNA encodi |
| 9 | 15.4 | 90.6 | 471 | E12010 | E12010 cDNA encodi |
| 10 | 15.4 | 90.6 | 471 | ASIGIF | Y09278 Artificial |
| 11 | 15.4 | 90.6 | 572 | MM06244 | U06244 Mus muscul |

```

12 15 4 90.6 866 94 MUSGIRPP
13 15 4 90.6 35143 90 AP0000229
14 15 4 90.6 59955 77 AC0800229
15 15 4 90.6 95155 88 AC078950
16 15 4 90.6 100000 90 AP000144
17 15 4 90.6 100634 90 AP001594
18 15 4 90.6 123631 91 HS22F01
19 15 4 90.6 202103 87 AC010176
20 15 4 90.6 340000 90 AP001695
21 15 4 84.7 355 54 G01707
22 15 4 84.7 710 94 MMBEGFL4
23 15 4 963 53 CNS07GZU
24 15 4 997 54 CNS07CNO
25 15 4 84.7 1084 53 CNS07OT9
26 15 4 84.7 1188 5 AF318610
27 15 4 84.7 1718 12 AF046922
28 15 4 84.7 1971 4 AB035057
29 15 4 84.7 2094 4 AB035066
30 15 4 84.7 2178 94 MSHBEGF04
31 15 4 84.7 2237 4 AB035069
32 15 4 84.7 2266 4 AB035062
33 15 4 84.7 2282 4 AB035061
34 15 4 84.7 2305 4 AB035065
35 15 4 84.7 2919 12 AF024633
36 15 4 84.7 3172 12 AF015753
37 15 4 84.7 3522 4 AB035058
38 15 4 84.7 4438 8 AF085184
39 15 4 84.7 19552 14 RICPPA
40 15 4 84.7 20128 13 AF155773
41 15 4 84.7 21077 5 CEK10C8
42 15 4 84.7 21644 92 HSH2A1KLT
43 15 4 84.7 22092 4 AC006637
44 15 4 84.7 24361 5 CELR02C2
45 15 4 84.7 27748 5 CELH34C03

```

ALIGNMENTS

```

RESULT 1
LOCUS AR072047 17 bp DNA
DEFINITION Sequence 5 from patent US 5912324.
ACCESSION AR072047
VERSION AR072047.1 GI:7222935
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 17)
AUTHORS Okamura,H., Tanimoto,T., Torioge,K., Kunikata,T., Taniguchi,M.,
Kohno,K. and Kurimoto,M.
TITLE Interferon-gamma (IFN-gamma.) inducing factor (IGIF, IL-18)
JOURNAL Patent: US 5912324-A 5 15-JUN-1999;
FEATURES
SOURCE Location/Qualifiers
1..17
/organism="unknown"
ORIGIN
4 a 2 c 4 g 3 t 4 others

```

```

Query Match 90.6%; Score 15.4; DB 9; Length 17;
Best Local Similarity 100.0%; Pred.No.3.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TTYGARGARATGATCC 17
1 TTYGARGARATGATCC 17
DB 1 TTYGARGARATGATCC 17
RESULT 2
AR072044

```

```

LOCUS AR072044 471 bp DNA
DEFINITION Sequence 1 from patent US 5912324.
ACCESSION AR072044
VERSION AR072044.1 GI:7222932
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 471)
AUTHORS Okamura,H., Tanimoto,T., Torioge,K., Kunikata,T., Taniguchi,M.,
Kohno,K. and Kurimoto,M.
TITLE Interferon-gamma (IFN-gamma.) inducing factor (IGIF, IL-18)
JOURNAL Patent: US 5912324-A 1 15-JUN-1999;
FEATURES
SOURCE Location/Qualifiers
1..471
/organism="unknown"
BASE COUNT 162 a 91 c 92 g 125 t 1 others
ORIGIN

```

```

Query Match 90.6%; Score 15.4; DB 9; Length 471;
Best Local Similarity 76.5%; Pred.No.4.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 TTYGARGARATGATCC 17
DB 244 TTYGARGARATGATCC 260
RESULT 3
LOCUS E13264 471 bp DNA
DEFINITION Mouse cDNA encoding a protein that induces to produce
interferon-gamma.
ACCESSION E13264
VERSION E13264.1 GI:3252069
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
REFERENCE 1 (bases 1 to 471)
AUTHORS Torioge,K., Tanimoto,T., Fukuda,S. and Kurimoto,M.
TITLE AGENT FOR SENSITIVE DISEASE
JOURNAL Patent: JP 1997157180-A 2 17-JUN-1997;
COMMENT HAYASHIBARA BIOCHEM LAB INC
OS Mus sp. (mouse)
PN JP 1997157180-A/2
PD 17-JUN-1997
PR 24-JAN-1996 JP 1996028722
PR 10-MAR-1995 JP 95P 78357, 29-SEP-1995 JP 95P 274988, PR
04-OCT-1995 JP 95P 279906
PI TORIOGE KAKUJI, TANIMOTO TADAO, FUKUDA SHIGETSU, PI
KURIMOTO MASASHI
PC A61K38/00,A61K38/00,A61K38/00,A61K38/00,C07K14/52,
PC C07K14/54,
PC C07K14/55;
CC strandedness: Double;
CC topology: Linear;
CC feature is identified by similarity;
FH key Location/Qualifiers
FT source 1..471
FT /organism="Mus sp."
FT /tisue="liver"
FT mat_peptide 1..471
FT /product="interferon-gamma inducer protein"
FT location/Qualifiers
1..471
/organism="Mus sp."
/db_xref="taxon:10095"
BASE COUNT 162 a 91 c 92 g 125 t 1 others

```

ORIGIN

Query Match 90.6%; Score 15.4; DB 10; Length 471;
 Best Local Similarity 76.5%; Pred. No. 4.5e+02;
 Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAGACC 17
 11:11:11:11:11:11
 Db 244 TTTGAGGAATGATGATCC 260

RESULT 4

LOCUS E14257 471 bp DNA PAT 28-JUL-1999
 DEFINITION CDNA encoding mouse interferon gamma-inducing factor.
 ACCESSION E14257
 VERSION E14257.1 GI:5708940
 KEYWORDS JP 1997289896-A/1.
 SOURCE Mus sp.
 ORGANISM Mus sp.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Akita,K., Nukada,Y., Fujii,M., Tanimoto,T. and Kurimoto,M.
 TITLE PROTEIN FOR INDUCING PRODUCTION OF INTERFERON-GAMMA IN
 IMMUNOCOMPETENT CELL
 JOURNAL Patent: JP 1997289896-A 1 11-NOV-1997;
 HAYASHIBARA BIOCHEM LAB INC

COMMENT

OS Mus sp. (mouse)
 PN JP 1997289896-A/1
 PD 11-NOV-1997
 PE 20-SEP-1996 JP 1996269105
 PR 26-SEP-1995 JP 95P 270725, 29-FEB-1996 JP 96P 67434 PI
 AKITA KENJI, NUKADA YOSHIYUKI, FUJII MITSUYUKI, TANIMOTO TADAO, PI
 KURIMOTO MASASHI
 PC C12P21/02,A61K38/00,A61K38/00,A61K38/00,A61K38/00, PC
 A61K38/00,
 PC A61K38/00,A61K38/00,C07K14/47,(C12P21/02,C12R1:91); CC
 strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 FH Key Location/Qualifiers
 FH source 1..471 /organism='Mus sp.'
 FT /tissue_type='liver'
 FT mat_deptide 1..471
 FT /product='mouse interferon gamma-inducing factor'

FEATURES
 source Location/Qualifiers
 1..471 /organism='Mus sp.'
 /db_xref='taxon:10095'

BASE COUNT 162 a 91 c 92 g 125 t 1 others
 ORIGIN

Query Match 90.6%; Score 15.4; DB 10; Length 471;
 Best Local Similarity 76.5%; Pred. No. 4.5e+02;
 Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAGACC 17
 11:11:11:11:11:11
 Db 244 TTTGAGGAATGATGATCC 260

RESULT 5
 LOCUS E14760 471 bp DNA PAT 28-JUL-1999
 DEFINITION cDNA encoding polypeptide which induces interferon gamma production
 HIGIF.

ACCESSION E14760
 VERSION E14760.1 GI:5709443
 KEYWORDS JP 1998007699-A/2.
 SOURCE Mus sp.
 ORGANISM Mus sp.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Ushio,S., Torigoe,K., Tanimoto,T., Okamura,H. and Kurimoto,M.
 TITLE 1 (bases 1 to 471)
 JOURNAL POLYPEPTIDE INDUCING PRODUCTION OF INTERFERON-GAMMA
 Patent: JP 1998007699-A 2 13-JAN-1998;
 HAYASHIBARA BIOCHEM LAB INC

COMMENT

OS Mus sp. (mouse)
 PN JP 1998007699-A/2
 PD 13-JAN-1998
 PE 18-SEP-1995 JP 1997058547
 PR 15-NOV-1994 JP 94P 304203
 PI USHIO SHINPEI, TORIGOE KAKUJI, TANIMOTO TADAO, OKAMURA HARUKI,
 PI KURIMOTO MASASHI
 PC C07K14/52,C07H21/04,C12N1/21,C12N15/09,C12P21/02//A61K38/00,
 PC (C12N1/21,
 PC C12R1:19),(C12P21/02,C12R1:19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 FH Key Location/Qualifiers
 FH source 1..471 /organism='Mus sp.'
 FT /tissue_type='liver'
 FT Location/Qualifiers
 1..471 /organism='Mus sp.'
 /db_xref='taxon:10095'

BASE COUNT 162 a 91 c 92 g 125 t 1 others
 ORIGIN

Query Match 90.6%; Score 15.4; DB 10; Length 471;
 Best Local Similarity 76.5%; Pred. No. 4.5e+02;
 Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAGACC 17
 11:11:11:11:11:11
 Db 244 TTTGAGGAATGATGATCC 260

RESULT 6

LOCUS E17139 471 bp DNA PAT 28-JUL-1999
 DEFINITION Murine mRNA for interleukin-18 (IL-18).
 ACCESSION E17139
 VERSION E17139.1 GI:5711822
 KEYWORDS JP 1998236974-A/5.
 SOURCE Mus sp.
 ORGANISM Mus sp.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Mathew,T.G., Nicholl,J.H., Udagawa,N. and Kurimoto,M.
 TITLE OSTEOCLAST-FORMATION INHIBITOR
 JOURNAL Patent: JP 1998236974-A 5 08-SEP-1998;
 HAYASHIBARA BIOCHEM LAB INC

COMMENT

OS Mus sp. (mouse)
 PN JP 1998236974-A/5
 PD 08-SEP-1998
 PE 25-FEB-1997 JP 1997055468
 PI MATTHEW TODD GALSPIE, NICHOLL JOY HOOMOOD, UDAGAWA NOBUYUKI,
 PI KURIMOTO MASASHI
 PC A61K38/00,A61K38/00//C07K14/54,C12N15/09;
 CC strandedness: Double;
 CC topology: Linear;

| | Key | Location/Qualifiers |
|--------------------------|---|---------------------------------|
| FH | | |
| FH | source | 1..471 |
| FT | | /organism='Mus sp.' |
| FT | | /tissue_type='liver' |
| FT | mat_peptide | 1..471 |
| FT | | /product='IL-18' |
| FEATURES | | |
| source | | location/Qualifiers |
| | | 1..471 |
| | | /organism='Mus sp.' |
| | | /db_xref='taxon:10095' |
| BASE COUNT | 162 a | 91 c 92 g 126 t |
| ORIGIN | | |
| Query Match | 90.6%; Score 15.4; DB 10; | Length 471; |
| Best Local Similarity | 76.5%; | Pred. No. 4.5e+02; |
| Matches 13; Conservative | 4; Mismatches 0; | Indels 0; Gaps 0; |
| OY | 1 TTTGANGARATGAYCC 17 | |
| | ::: ::: ::: :: | |
| DB | 244 TTTGAGGAATGCATCC 260 | |
| RESULT | 7 | |
| EI0609 | standard; RNA; ROD; 471 BP. | |
| XX | EI0609 | |
| AC | EI0609; | |
| SV | EI0609.1 | |
| XX | | |
| XX | 08-OCT-1997 (Rel. 52, Created) | |
| DT | 02-SEP-2000 (Rel. 65, Last updated, Version 2) | |
| DE | Mouse cDNA encoding a protein involved in interferon-gamma production. | |
| KM | JF 1996027189-A/1. | |
| OS | Mus sp. | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; | |
| OC | Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | |
| RN | [1] | |
| RP | 1-471 | |
| RA | Okamura H., Tanimoto T., Torigoe K., Kurimoto M.; | |
| RT | "PROTEIN INDUCING PRODUCTION OF INTERFERON-GAMMA"; | |
| RL | Patent number JP1996027189-A/1, 30-JAN-1996. | |
| HL | HAYASHIBARA BIOCHEM LAB INC. | |
| XX | | |
| OS | Mus sp. (mouse) | |
| PN | JF 1996027189-A/1 | |
| PD | 30-JAN-1996 | |
| CC | PF 14-JUL-1994 JF 1994184162 | |
| CC | PI OKAMURA HARUKI, TANIMOTO TADAO, TORIGOE KAKUJI, | |
| CC | KURIMOTO MASASHI | |
| CC | PC C07K14/52,A6K38/00,A61K38/00,C12N1/21,C12N15/09, | |
| CC | C12P21/02//C07K14/57; | |
| CC | PC C12P21/02//C07K14/57; | |
| CC | strandedness: Double; | |
| CC | topology: Linear; | |
| CC | key | location/Qualifiers |
| FH | | |
| CC | source | 1..471 |
| CC | | /organism='Mus sp.' |
| CC | | /tissue_type='liver' |
| CC | mat_peptide | 1..471 |
| CC | | /product='a protein involved in |
| CC | | interferon-gamma |
| CC | | production" |
| FT | | |
| XX | | |
| XX | key | Location/Qualifiers |
| FH | | |
| FT | source | 1..471 |

| | |
|---|---|
| FT | /db_xref="taxon:10095" |
| FT | /organism="Mus sp." |
| SO | Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other; |
| Query Match | 90.6%; Score 15.4; DB 45; Length 471; |
| Best Local Similarity | 76.5%; Pred. No. 4.5e+02; |
| Matches 13; Conservative | 4; Mismatches 0; Indels 0; Gaps 0; |
| Oy | 1 TTYGARGARATGATCC 17 11:11:11:11:11:11 |
| Db | 244 TTYGAGGAATGATGCC 260 |
| RESULT 8 | |
| ID E11744 | standard; RNA; ROD; 471 BP. |
| XX E11744; | |
| AC E11744; | |
| XX E11744.1 | |
| SV E11744.1 | |
| XX 08-OCT-1997 (Rel. 52, Created) | |
| DT 02-SEP-2000 (Rel. 65, Last updated, Version 2) | |
| XX cDNA encoding polypeptide which induce mouse interferon-gamma product. | |
| DE JP 1996193098-A/1. | |
| XX Mus musculus (house mouse) | |
| OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia | |
| OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | |
| XX [1] | |
| RP 1-471 | Ushio S., Torioge K., Tanimoto T., Okamura H., Kurimoto M.; |
| RA "POLYPEPTIDE FOR INDUCING PRODUCTION OF INTERFERON-GAMMA"; | |
| RT Patent number JP1996193098-A/1. 30-JUL-1996. | |
| RL HAYASHIBARA BIOCHEM LAB INC. | |
| XX OS Mus musculus (mouse) | |
| CC PN JP 1996193098-A/1 | |
| CC PD 30-JUL-1996 | |
| CC PF 18-SEP-1995 JP 1995262062 | |
| CC PF 15-NOV-1994 JP 946 304203 | |
| CC PI USHIO SHINPEI, TORIOGE KAZUJI, TANIMOTO TADAO, OKAMURA HARUKI, | |
| CC PI KURIMOTO MASASHI | |
| CC PC C07K14/52,C07H21/04,C12N1/21,C12N15/09,C12P21/00/A6138/00, | |
| CC PC C07K7/06, | |
| CC PC C07K7/08,(C12N1/21,C12N1.19),(C12P21/00,C12R1.19); | |
| CC CC strandedness: Double; | |
| CC CC topology: Linear; | |
| CC CC hypothetical: No; | |
| CC CC anti-sense: No; | |
| CC FH Key | Location/Qualifiers |
| CC FT source | 1..471 |
| CC FT /organism="Mus musculus" | |
| CC FT /tissue-type="liver" | |
| CC XX | |
| FH Key | Location/Qualifiers |
| FH source | 1..471 |
| FT /db_xref="taxon:10090" | |
| FT /organism="Mus musculus" | |
| SO Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other; | |
| Query Match | 90.6%; Score 15.4; DB 45; Length 471; |
| Best Local Similarity | 76.5%; Pred. No. 4.5e+02; |
| Matches 13; Conservative | 4; Mismatches 0; Indels 0; Gaps 0; |

QY 1 TTYGARGARATGAYCC 17
 ||:||||:||||:|
 Db 244 TTTGAGGAATGATGCC 260

RESULT 9
 E12010 standard; RNA; ROD: 471 BP.
 XX E12010:
 4 AC E12010:
 XX E12010.1
 XX 07-OCT-1997 (Rel. 52, Created)
 DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
 XX
 DE cDNA encoding mouse polypeptide which introduce interferon-gamma product in
 DE immunocompetent cell.
 XX
 XX JP 1996231598-A/2.

OS Mus musculus (house mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 XX
 XX [1]
 RN 1-471
 RA Kunikata T., Taniguchi M., Kono K., Kurimoto M.;
 RT "MONOCLONAL ANTIBODY";
 RL Patent number JP1996231598-A/2, 10-SEP-1996.
 RL HAYASHIBARA BIOCHEM LAB INC.
 XX

OS Mus musculus (mouse)
 CC PN JP 1996231598-A/2
 CC PD 10-SEP-1996
 CC PF 23-FEB-1995 JP 1995058240
 CC PI KUNIKATA TOSHIO, TANIGUCHI MUTSUO, KONO KEIZO,
 CC PI KURIMOTO MASASHI
 CC PC C07K16/24, C07K1/16, C07K1/18, C07K1/22, C07K1/26, C07K1/30,
 CC PC C07K1/34, C12N5/10,
 CC PC C12N15/02, C12P21/08, G01N33/53, G01N33/577//A61K38/21, A61K39/395,
 CC PC (C12P21/08,
 CC PC C12K1/91);
 CC PC strandedness: Double;
 CC CC topology: Linear;
 CC CC hypothetical: No;
 CC CC anti-sense: No;
 CC CC Key Location/Qualifiers
 CC FH source 1..471
 CC FT /organism="Mus musculus"
 CC FT /tissue_type="liver"
 XX
 XX
 FH Key Location/Qualifiers
 FH source 1..471
 FT /db_xref="taxon:10090"
 FT /organism="Mus musculus"
 XX
 SO Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other;

Query Match 90.6%; Score 15.4; DB 45; Length 471;
 Best Local Similarity 76.5%; Pred. No. 4.5e+02;
 Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTYGARGARATGAYCC 17
 ||:||||:||||:|
 Db 244 TTTGAGGAATGATGCC 260

RESULT 10

ASIGIF 471 bp DNA SYN 08-NOV-1996
 LOCUS Artificial sequence DNA for cytokine interferon-gamma inducing
 DEFINITION factor (IGIF).
 ACCESSION Y09278.1 GI:1666283
 VERSION Y09278.1
 KEYWORDS cytokine; interferon-gamma inducing factor (IGIF).
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 471)
 AUTHORS Schluesener, H.J.
 JOURNAL unpublished
 REFERENCE 2 (bases 1 to 471)
 AUTHORS Schluesener, H.J.
 TITLE Direct Submission
 JOURNAL Submitted (21-OCT-1996) H.J. Schluesener, Inst.f. Hirnforschung,
 Calwer Str. 3, D-77076 Tübingen, FRG
 FEATURES
 source Location/Qualifiers
 1..471
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 <1..>471
 /codon_start=1
 /transl_table=1
 /product="mouse cytokine interferon-gamma inducing factor"
 /protein_id="CAA0473.1"
 /db_xref="GI:1666284"
 /translation="NFGRLHCTAVIRINDQVLFVDKROPVPEMDTIDQASSEPOT
 RLITYMADSEVGLAVTLVYKDSKMSLCKNLTISFEEDPDPNDIDQSLIFPQ
 KRVPGNHMEESLYEGHFLACKEEDAFRLIKKRDENGDSVMTLTNLMOS"

BASE COUNT 148 a 100 c 130 g 93 t

Query Match 90.6%; Score 15.4; DB 56; Length 471;
 Best Local Similarity 76.5%; Pred. No. 4.5e+02;
 Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGARATGAYCC 17
 ||:||||:||||:|
 Db 244 TTTGAGGAATGATGCC 260

RESULT 11
 MMU66244 572 bp mRNA ROD 18-MAR-1997
 LOCUS Mus musculus interferon-gamma inducing factor mRNA, partial cds.
 DEFINITION U66244
 ACCESSION U66244
 VERSION U66244.1 GI:1561735
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 572)
 AUTHORS Rothe, H., Jenkins, N.A., Copeland, N.G. and Kolb, H.
 TITLE Active stage of autoimmune diabetes is associated with the
 expression of a novel cytokine, IGIF, which is located near Idd2
 J. Clin. Invest. 99 (3), 469-474 (1997)
 JOURNAL 97174346
 MEDLINE
 REFERENCE 2 (bases 1 to 572)
 AUTHORS Rothe, H., Copeland, N.G. and Kolb, H.
 TITLE Direct Submission
 JOURNAL Submitted (06-AUG-1996) Diabetes Research Institute, Auf'm
 Hennekamp 65, Dueseldorf 40225, Germany
 FEATURES
 source Location/Qualifiers
 1..572
 /organism="Mus musculus"
 /strain="non obese diabetic (NOD)"
 /db_xref="taxon:10090"
 /chromosome="9"
 /tissue_type="pancreas"

```

CDS
1. >572
/note="IGIF"
/codon_start=1
/product="interferon gamma inducing factor"
/protein_id="VAB49753.1"
/db_xref="GI:1561736"
/translation="MAAMESDCVNFKEKMFIDNTLYFPEENGDESNGFRLHCTT
AVIRINDOVLVDKRPVDEMDIDDSASBPOTRLITYKDESEVGLAVTLVKD
SKMSTLCKNRKIISFEEDMDPENIDIOSDILFFORVPHNKEFESSLYEGHFLAC
OKEDDAFKLLKKRDKDNGSKSVISLTNLHQ"
BASE COUNT      196 a      111 c      113 g      152 t
ORIGIN

Query Match      90.6%; Score 15.4; DB 94; Length 572;
Best Local Similarity 76.5%; Pred. No. 4.6e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARGATGCATCC 17
Db 349 TTYGAGGAATGATCC 365

RESULT 12
LOCUS      MUSGIFPP      866 bp      mRNA      ROD      10-FEB-1999
DEFINITION Mouse mRNA for IGIF precursor polypeptide, complete cds.
ACCESSION D49949.1 GI:1064822
VERSION D49949.1
KEYWORDS IGIF precursor polypeptide.
SOURCE Mus musculus liver cDNA to mRNA, clone pMUGF37B-5.
ORGANISM Mus musculus
Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 866)
Okamura, H.
Direct Submission
Submitted (29-MAR-1995) to the DDBJ/EMBL/Genbank databases. Haruki
Okamura, Hyogo College of Medicine, Department of Bacteriology;
Mikogawa 1-1, Nishinomiya, Hyogo 653, Japan (Tel:0798-45-6111);
2 (bases 1 to 866)
Okamura, H., Tsutui, H., Komatsu, T., Yutsudo, M., Hakura, A.,
Tanimoto, T., Torigoe, K., Okura, T., Nukada, S., Hattori, K., Akita, K.,
Namba, M., Tanabe, F., Konishi, K., Fukuda, S. and Kurimoto, M.
Cloning of a new cytokine that induces IFN-gamma production by T
cells
JOURNAL Nature 378 (6552), 88-91 (1995)
MEDLINE 96061009
FEATURES
source      Location/Qualifiers
1..866
/organism="Mus musculus"
/db_xref="taxon:10090"
/locus_type="liver"
1..866
/note="interferon-gamma inducing factor (IGIF) mRNA"
165..743
/codon_start=1
/product="IGIF precursor polypeptide"
/protein_id="BA08705.1"
/db_xref="GI:1064823"
/translation="MAAMESDCVNFKEKMFIDNTLYFPEENGDESNGFRLHCTT
AVIRINDOVLVDKRPVDEMDIDDSASBPOTRLITYKDESEVGLAVTLVKD
SKMSTLCKNRKIISFEEDMDPENIDIOSDILFFORVPHNKEFESSLYEGHFLAC
OKEDDAFKLLKKRDKDNGSKSVISLTNLHQ"
BASE COUNT      262 a      187 c      187 g      230 t
ORIGIN

polyA_site
BASE COUNT      262 a      187 c      187 g      230 t
ORIGIN

Query Match      90.6%; Score 15.4; DB 94; Length 866;
Best Local Similarity 76.5%; Pred. No. 4.7e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARGATGCATCC 17
Db 513 TTYGAGGAATGATCC 529

RESULT 13
LOCUS      AP000229/c      35143 bp      DNA      prt      17-MAR-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q21.1-q21.2, clone:11539,
L156 A56 region, complete sequence.
ACCESSION AP000229.2 GI:7262565
VERSION AP000229.2
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:11539.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 35143)
Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Hattori, W., Ishii, K., Tokoi, Y. and Sakaki, Y.
Fujiyama, A., Ishii, K., Tokoi, Y. and Sakaki, Y.
Homo sapiens 35,143 genomic DNA of 21q21.1-q21.2
Published only in Database (1999) In press
2 (bases 1 to 35143)
Hattori, W., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Iida, T., Tokoi, Y. and Sakaki, Y.
Direct Submission
Submitted (13-MAR-1999) to the DDBJ/EMBL/Genbank databases.
Masahito Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagami-ku, Kanagawa 228-8555, Japan
Tel:81-12-778-9923. Fax:81-12-778-9924.
URL: http://npg.gsc.riken.go.jp/
O1 Mar 17, 2000 this sequence version replaced gi:4835598.
Sequence updated (15-Mar-2000)
COMMENT The sequencing project is supported by Japan Science Technology
Corporation (JST) and The Institute of Physical and Chemical
Research (RIKEN)
FEATURES
source      Location/Qualifiers
1..35143
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="11539"
/ma="21q21.1-q21.2"
BASE COUNT      10208 a      7036 c      7565 g      10334 t
ORIGIN

Query Match      90.6%; Score 15.4; DB 90; Length 35143;
Best Local Similarity 76.5%; Pred. No. 6.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARGATGCATCC 17
Db 25608 TTYGAGGAATGATCC 25592

RESULT 14
LOCUS      AC080029      5995 bp      DNA      HTG      25-OCT-2000
DEFINITION Homo sapiens chromosome 4 clone CTD-2353K2 map 4, LOW-PASS SEQUENCE
ACCESSION AC080029.1 GI:10280784
VERSION AC080029.1
KEYWORDS HTG, HTGS_PHASE0.
SOURCE Human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (base 1 to 5995)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

```

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 4, clone CTD-2353K2
Unpublished
2 (bases 1 to 59955)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,
Bounguelter, B., Brown, A., Burkett, G., Campoliano, A., Castile, A.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeAtellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
FitzHugh, W., Gage, D., Galagan, J., Gardina, S., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lacombe, K.,
Lamares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,
McPeckers, R., Meldrim, J., Menus, L., Mihova, T., Mlenka, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogov, P., Rotman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sounguez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L10743
Center clone name: 2353_K_2

* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 645: contig of 645 bp in length
* 646 745: gap of 100 bp
* 746 1464: contig of 719 bp in length
* 1465 1564: gap of 100 bp
* 1565 2315: contig of 751 bp in length
* 2316 2415: gap of 100 bp
* 2416 3059: contig of 644 bp in length
* 3060 3159: gap of 100 bp
* 3160 3877: contig of 718 bp in length
* 3878 3977: gap of 100 bp
* 3978 4722: contig of 745 bp in length
* 4723 4822: gap of 100 bp
* 4823 5579: contig of 757 bp in length
* 5580 5679: gap of 100 bp
* 5680 6315: contig of 636 bp in length
* 6316 6415: gap of 100 bp
* 6416 7280: contig of 865 bp in length
* 7281 7380: gap of 100 bp
* 7381 8134: contig of 754 bp in length
* 8135 8234: gap of 100 bp
* 8235 8980: contig of 746 bp in length
* 8981 9080: gap of 100 bp
* 9081 9763: contig of 683 bp in length
* 9764 9863: gap of 100 bp

* 9864 10479: contig of 616 bp in length
* 10480 10579: gap of 100 bp
* 10580 11203: contig of 624 bp in length
* 11204 11303: gap of 100 bp
* 11304 12043: contig of 740 bp in length
* 12044 12143: gap of 100 bp
* 12144 12846: contig of 703 bp in length
* 12847 12946: gap of 100 bp
* 12947 13678: contig of 732 bp in length
* 13679 13778: gap of 100 bp
* 13779 14542: contig of 764 bp in length
* 14543 14642: gap of 100 bp
* 14643 15366: contig of 724 bp in length
* 15367 15466: gap of 100 bp
* 15467 16160: contig of 694 bp in length
* 16161 16260: gap of 100 bp
* 16261 17065: contig of 805 bp in length
* 17066 17165: gap of 100 bp
* 17166 17914: contig of 749 bp in length
* 17915 18014: gap of 100 bp
* 18015 18715: contig of 701 bp in length
* 18716 18815: gap of 100 bp
* 18816 19436: contig of 621 bp in length
* 19437 19536: gap of 100 bp
* 19537 20176: contig of 640 bp in length
* 20177 20276: gap of 100 bp
* 20277 21013: contig of 737 bp in length
* 21014 21113: gap of 100 bp
* 21114 21872: contig of 759 bp in length
* 21873 21972: gap of 100 bp
* 21973 22634: contig of 662 bp in length
* 22635 22734: gap of 100 bp
* 22735 23340: contig of 606 bp in length
* 23341 23440: gap of 100 bp
* 23441 24159: contig of 719 bp in length
* 24160 24259: gap of 100 bp
* 24260 24993: contig of 734 bp in length
* 24994 25093: gap of 100 bp
* 25094 25806: contig of 713 bp in length
* 25807 25906: gap of 100 bp
* 25907 26519: contig of 613 bp in length
* 26520 26619: gap of 100 bp
* 26620 27253: contig of 634 bp in length
* 27254 27353: gap of 100 bp
* 27354 28058: contig of 705 bp in length
* 28059 28158: gap of 100 bp
* 28159 28834: contig of 676 bp in length
* 28835 28934: gap of 100 bp
* 28935 29594: contig of 660 bp in length
* 29595 29694: gap of 100 bp
* 29695 30458: contig of 764 bp in length
* 30459 30558: gap of 100 bp
* 30559 31323: contig of 765 bp in length
* 31324 31423: gap of 100 bp
* 31424 32170: contig of 747 bp in length
* 32171 32270: gap of 100 bp
* 32271 33037: contig of 767 bp in length
* 33038 33137: gap of 100 bp
* 33138 33845: contig of 708 bp in length
* 33846 33945: gap of 100 bp
* 33946 34637: contig of 692 bp in length
* 34638 34737: gap of 100 bp
* 34738 35427: contig of 690 bp in length
* 35428 35527: gap of 100 bp
* 35528 36173: contig of 646 bp in length
* 36174 36273: gap of 100 bp
* 36274 36921: contig of 648 bp in length
* 36922 37021: gap of 100 bp
* 37022 37736: contig of 715 bp in length
* 37737 37836: gap of 100 bp
* 37837 38624: contig of 788 bp in length
* 38625 38724: gap of 100 bp
* 38725 39480: contig of 756 bp in length

reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Contig length: 173652
Phrap values in estimate: 170386
Average error rate (BCM-Phrap estimate): 0.000194024
Fraction of Phrap values less than 40 : 0.0280363
Number of consensus changing edits: 1401
Number of N's in consensus : 0
```

```
----- Consensus changing edits -----
Position Original+Context Edited+Context
1 . . . . . (n)nnnnnnnnnn . . . . . (t)aggt
2 . . . . . n(n)nnnnnnnnnn . . . . . ta(g)gtgcaag
3 . . . . . .nn(n)nnnnnnnnnn . . . . . tagg(t)tgcaagaaac
4 . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
5 . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
6 . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
7 . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
8 . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
9 . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
10 . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
11 . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
12 . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
13 . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
14 . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
15 . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
16 . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
17 . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
18 . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
19 . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
20 . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
21 . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
22 . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
23 . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
24 . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
25 . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
26 . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
27 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
28 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
29 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
30 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
31 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
32 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
33 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
34 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
35 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
36 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
37 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
38 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
39 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
40 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
41 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
42 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
43 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
44 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
45 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
46 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
47 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
48 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggt(t)gcaagaaacn
```

```
49 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . taggtatggt(a)atgtgcaag
50 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . ttatgtgta(a)atgtgcaag
51 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
52 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
53 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
54 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
55 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
56 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
57 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
58 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
59 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
60 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
61 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
62 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
63 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
64 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
65 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
66 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
67 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
68 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
69 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
70 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
71 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
72 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
73 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
74 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
75 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
76 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
77 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
78 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
79 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
80 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
81 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
82 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
83 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .nnnn(n)nnnnnnnnnn . . . . . tatgtgtta(a)atgtgcaag
```

Query Match 90.6%; Score 15.4; DB 88; Length 95155;

Best Local Similarity 76.5%; Pred. No. 7.1e+02; Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGATGATCC 17

DB 69860 TTYGAAGATGATCC 69844

RESULT 16

AP000144/c 100000 bp DNA 08-JAN-2000

LOCUS Homo sapiens genomic DNA, chromosome 21q21.2, LL56-APP region,

DEFINITION clone B2291C14-R44F3, segment 9/10, complete sequence.

ACCESSION AP000144

VERSION AP000144.1 GI:4827110

KEYWORDS HGC.

SOURCE Homo sapiens DNA.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 100000) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,

TITLE Fujiyama, A., Yada, T., Totoki, Y., and Sakaki, Y.

JOURNAL CLONE RANGE: B2291C14-R44F3

REFERENCE 2 (bases 1 to 100000) Published Only in Database (1999) In press

AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,

TITLE Fujiyama, A., Yada, T., Totoki, Y., and Sakaki, Y.

JOURNAL Direct Submission

Submitted (10-MAY-1999) to the DDBJ/EMBL/Genbank databases.

Maehara Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1

Kitasato, Sagamihara, Kanagawa 228-8555, Japan

(E-mail: hattori@gscc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>,

tel: 81-42-778-9923, fax: 81-42-778-9924)


```

FEATURES
Source
> 'tRNA Scan': tRNAscan-SE (Lowe & Eddy), Vers. 1.11.
Location/Qualifiers
1..123631
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="LINC021F0122"
/map="21q21"
606..928
/misc_feature
/note="match: EST Z39889"
complement(620..787)
exon
/note="GRAIL, score = 54%, comment = good"
complement(620..739)
exon
/note="GENSCAN prediction, score = 0.29"
731..882
exon
/note="WZEF prediction, score = 0.535"
1007..1027
satellite
/note="AAAT repeat"
complement(1073..1207)
exon
/note="GENSCAN prediction, score = 3.34"
complement(1075..1165)
exon
/note="GRAIL, score = 62%, comment = good"
complement(1089..1207)
exon
/note="WZEF prediction, score = 0.884"
complement(1174..1491)
misc_feature
/note="match: EST Z43821"
complement(1304..1353)
exon
/note="WZEF prediction, score = 0.773"
1587..1627
exon
/note="GRAIL, score = 62%, comment = good shadow; WZEF
prediction, score = 0.886"
(1842..2055)..(2141..2222)
misc_feature
/note="match: ESTs Z17821 L25498"
complement(1978..2021)
exon
/note="XPOND prediction, score = 0.412"
complement(2133..2214)
exon
/note="XPOND prediction, score = 0.367"
2191..2490
exon
/note="GeneFinder prediction"
complement(2478..2615)
exon
/note="GeneFinder prediction"
2486..4115
misc_feature
/note="CpG_island (%GC=69.4, o/e=0.89, #CpGs=168)"
2366..2933
exon
/note="GeneFinder prediction"
complement(2769..2880)
exon
/note="GeneFinder prediction"
2811..2904
exon
/note="XPOND prediction, score = 0.326"
3009..3024
satellite
/note="CTC repeat"
3029..3331
exon
/note="GeneFinder prediction"
complement(3213..3386)
exon
/note="GeneFinder prediction"
3321..3333
exon
/note="XPOND prediction, score = 0.205"
3423..3961
exon
/note="XPOND prediction, score = 0.843"
3442..3965
exon
/note="GeneFinder prediction"
3470..3586
misc_feature
/note="match: EST AA931442"
complement((3471..3489)..(3597..3657))
misc_feature
/note="match: ESTs AA114221 AL042549 Z84022"
complement(3474..4166)
exon
/note="GeneFinder prediction"
complement(3474..3675)
exon
/note="GENSCAN prediction, score = 12.87"
complement(3474..3530)
exon
/note="GRAIL, score = 94%, comment = excellent"
complement(3894..4140)
exon
/note="GENSCAN prediction, score = 4.30"

```

```

exon
3973..4081
/note="XPOND prediction, score = 0.424"
complement(4737..4843)
exon
/note="GeneFinder prediction: GRail, score = 53%, comment
= good"
5074..5095
repeat_region
/note="IR1, 90% complementary to IR1' (5172..5193)"
/rpt_type=INVERTED
5172..5193
repeat_region
/note="IR1', 90% complementary to IR1 (5074..5095)"
/rpt_type=INVERTED
5371..5387
exon
/note="XPOND prediction, score = 0.221"
5422..5699
repeat_region
/note="91% identity: matches 1..277 of consensus"
/rpt_family="AluSx"
complement(5436..5700)
repeat_region
/note="85% identity: matches 165..424 of consensus"
/rpt_family="L1"
complement(5873..6117)
repeat_region
/note="91% identity: matches 245..489 of consensus"
/rpt_family="L1"
5873..6117
repeat_region
/note="96% identity: matches 37..280 of consensus"
/rpt_family="AluSb"
5889..6113
misc_feature
/note="CpG_island (%GC=60.4, o/e=0.90, #CpGs=18)"
6229..6340
exon
/note="XPOND prediction, score = 0.550"
complement(7110..7159)
misc_feature
/note="match: EST AW138710"
complement(7161..7395)
repeat_region
/note="100% identity: matches 1..235 of consensus"
/rpt_family="MLTB"
7223..7336
exon
/note="GRAIL, score = 59%, comment = good"
complement(7336..7630)
misc_feature
/note="match: EST AW138710"
complement(7474..7507)
misc_feature
/note="match: GSS A0895571"
complement(7663..7737)
exon
/note="WZEF prediction, score = 0.552"
complement(7828..7903)
repeat_region
/note="88% identity: matches 23..99 of consensus"
/rpt_family="MLTB"
complement(8134..8154)..(8183..8194))
misc_feature
/note="match: GSS A080344 AF057094"
8150..8194
misc_feature
/note="match: EST A1034376"
complement(8156..8185)
misc_feature
/note="match: EST AW002649"
8411..8469
exon
/note="XPOND prediction, score = 0.275"
complement(8565..8745)
misc_feature
/note="match: GSS B76378"
complement(8759..9069)
misc_feature
/note="match: GSS B76378"

```

```

Query Match          90.6%: Score 15.4; DB 91: Length 123631;
Best Local Similarity 76.5%: Pred. No. 7.3e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY      1 TTYGARGARGATGAYCC 17
DB 38138 TTTGAGGAATGATGCC 38122

```

```

RESULT 19
AC010176 202103 bp DNA PRI 19-SEP-2000
LOCUS Homo sapiens 12 BAC RP11-711K1 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC010176

```

VERSION AC010176.12 GI:10190751

KEYWORDS

SOURCE human

ORGANISM Homo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 202103)

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

Alibonick, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbra, J.,

Benton, J., Bimige, K., Blankenburg, K., Bonini, D., Bouck, J.,

Bowie, S., Brivett, M., Brown, E., Brown, M., Bryant, N.P., Bulay, C.,

Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C., Carron, T.F.,

Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, R.,

Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D.,

Cox, C., Coyle, M.D., Dalbourn, S.R., David, R., Davila, M.L., Davis, C.,

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,

Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,

Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C.,

Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D.,

Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J.,

Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W.,

Gunnar, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K.,

Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O.,

Howson, A., Hughes, M., Holmway, C., Hollins, B., Homsl, F.,

Howard, S., Huber, J., Hult, S., Hume, J., Ioshikhes, I., Jackson, L.E.,

Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,

Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,

Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C.,

Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W.,

Louise, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R.,

Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P.,

Meador, M., Mei, G., Merscher, S., Metzger, M., Miller, A., Miner, G.,

Miner, Z., Mitchell, T., Mohabhat, K., Montgomery, K.T., Morgan, M.,

Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N.,

Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenwo, S.,

Ogih, M., Okunou, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,

Peery, J., Peters, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,

Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,

Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shm, C.,

Shoshari, N., Slisson, I., Sodergren, E., Sonaike, T., Sparks, A.,

Stanley, H., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, A.,

Thomas, K., Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalón, D.,

Vinson, R., Wall, R., Wang, S., Ward, Moore, S., Warren, R.,

Washington, C., Wallington, S., Williams, G., Williamson, A.,

Wieczyk, R., Woden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J.,

Zorilla, S., Zuercher, R., and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 202103)

Worley, K.C.

Direct Submission

Submitted (15-SEP-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 202103)

Worley, K.C.

Direct Submission

Submitted (19-SEP-2000) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Sep 19 2000 this sequence version replaced gi:9887569.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> OR email gc-help@bcm.tmc.edu

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT

----- Summary Statistics -----
Contig length: 202103
Phrap values in estimate: 201669
Average error rate (BCM-phrap estimate): 9.77954e-05
Fraction of Phrap values less than 40: 0.0287005
Number of consensus changing edits: 29
Number of N's in consensus: 0

| Position | Original/Context | Edited/Context |
|----------|------------------------|------------------------|
| 11794 | taattcctt(n)catgcatc | taattcctt(n)catgcatc |
| 13909 | atgaacaa(n)ctcccatc | atgaacaa(n)ctcccatc |
| 39837 | agagggcaat(n)tatgacaa | agagggcaat(n)tatgacaa |
| 39928 | aaaggaac(n)catgaag | aaaggaac(n)catgaag |
| 45888 | cacttgatt(n)taaacacaa | cacttgatt(n)taaacacaa |
| 54727 | aaagaagat(n)caaaatc | aaagaagat(n)caaaatc |
| 54740 | aaatctg(n)cttccctc | aaatctg(n)cttccctc |
| 73838 | cttccccc(n)ctccatcc | cttccccc(n)ctccatcc |
| 73861 | cttccccc(n)ctccatcc | cttccccc(n)ctccatcc |
| 89489 | tacattgat(n)tgatcctaa | tacattgat(n)tgatcctaa |
| 89496 | atnigtatc(n)taaaattg | atnigtatc(n)taaaattg |
| 89499 | tgatcctaa(n)aaattgctg | tgatcctaa(n)aaattgctg |
| 90500 | tgatcctaa(n)aaattgctg | tgatcctaa(n)aaattgctg |
| 90503 | ttttattatg(n)acttgctga | ttttattatg(n)acttgctga |
| 93328 | tgtaataaa(n)caaaaagaa | tgtaataaa(n)caaaaagaa |
| 93335 | aaacaaaaa(n)gaaaaaaga | aaacaaaaa(n)gaaaaaaga |
| 125580 | atcgtcttg(n)taaatatt | atcgtcttg(n)taaatatt |
| 135997 | ctcttcacg(n)ccgcaacg | ctcttcacg(n)ccgcaacg |
| 139661 | gagccagat(c)ccgcaacg | gagccagat(c)ccgcaacg |
| 151709 | acatgaggt(n)gtcaattc | acatgaggt(n)gtcaattc |
| 199225 | agaaaggga(n)catlaaac | agaaaggga(n)catlaaac |
| 199339 | ctcccaacc(n)ttccctgac | ctcccaacc(n)ttccctgac |
| 199375 | ctcccaacc(n)ttccctgac | ctcccaacc(n)ttccctgac |
| 199378 | ctcccaacc(n)ttccctgac | ctcccaacc(n)ttccctgac |
| 199380 | ctcccaacc(n)ttccctgac | ctcccaacc(n)ttccctgac |
| 199383 | ctcccaacc(n)ttccctgac | ctcccaacc(n)ttccctgac |
| 200220 | gtcgcctta(n)ttcctgac | gtcgcctta(n)ttcctgac |
| 200366 | aaatacaaga(n)caagagggc | aaatacaaga(n)caagagggc |

----- Distribution of Quality < 40 Bases -----

| # | bases | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
|-------|-------|---|----|----|----|----|----|----|----|
| 10001 | | | | | | | | | * |
| 9001 | | | | | | | | | * |
| 8001 | | | | | | | | | * |
| 7001 | | | | | | | | | * |
| 6001 | | | | | | | | | * |
| 5001 | | | | | | | | | * |
| 4001 | | | | | | | | | * |
| 3001 | | | | | * | * | * | * | * |
| 2001 | | | | * | * | * | * | * | * |
| 1001 | | | | * | * | * | * | * | * |
| 01 | | * | * | * | * | * | * | * | * |

Phrap Value Range

Version: 1.01 gxf.
 Location/Qualifiers
 Source 1. 202103
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="12"
 /clone="RP11-71K1"
 /complement(1..189)
 /rpt_family="L1P"
 190..447
 /rpt_family="HSMAR2"
 1208..1494
 /rpt_family="AluDo"
 1944..2300
 /rpt_family="MLT1B"
 3134..4152
 /rpt_family="HSMAR2"
 4188..5034
 /rpt_family="LTR12"
 5035..5135
 /rpt_family="HSMAR2"
 5138..5296
 /rpt_family="HERVL40"
 5857..5891
 /rpt_family="(TTA)n"
 complement(5894..6183)
 /rpt_family="AluDo"
 6547..6622
 /rpt_family="HERVL40"
 6759..6947
 /rpt_family="LINE"
 7027..7173

Query Match 90.6%; Score 15.4; DB 87; Length 202103;
 Best Local Similarity 76.5%; Pred. No. 7.5e+02;
 Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGATGATGATCC 17
 11:11:11:11:11:11
 Db 1318 TTTGAGAGATGATGCC 1302

RESULT 20
 AP001695/c DNA PRI 30-MAY-2000
 LOCUS AP001695 340000 bp
 DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 39/105.
 ACCESSION AP001695 A1163240 BA000005
 VERSION AP001695.1 GI:7768704
 KEYWORDS
 SOURCE Homo sapiens DNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eulalia; Primates; Catarrhini; Homnidae; Homo.
 Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T.,
 REFERENCES
 1 (sites)
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 2 (bases 1 to 340000)
 Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T.,
 Park, H. S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, E.,
 Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K.,
 Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R.,
 Patterson, D., Reichwald, K., Rump, A., Schillabel, M., and Schudy, A.
 The DNA sequence of human chromosome 21. The chromosome 21 mapping
 and sequencing consortium
 Nature 405 (6784), 311-319 (2000)
 20289799

COMMENT

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Direct Submission
 Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The
 Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic
 Sciences Center, Human Genome Research Group * Institute of
 Molecular Biotechnology, Genome Analysis * Keio University School
 of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome
 Analysis * Max-Planck Institute for Molecular Genetics (addresses
 see below)
 On May 30, 2000 this sequence version replaced gi:7717293.
 The chromosome 21 mapping and sequencing consortium consisting of
 * RIKEN Genomic Sciences Center, Human Genome Research Group, *
 Sagami-hara 228-8555, Japan,
 * e-mail: hattori@gscl.riken.go.jp
 * URL: http://hgp.gsc.riken.go.jp/
 and
 * Institute of Molecular Biotechnology, Genome Analysis, *
 Beutenbergstrasse 11, D-07745 Jena, Germany,
 * e-mail: gscj-submit@genome.imb-jena.de
 * URL: http://genome.imb-jena.de/
 and
 * Keio University School of Medicine, Molecular Biology, * Tokyo
 160-8582, Japan,
 * e-mail: nshimizu@med.keio.ac.jp
 * URL: http://www.dmb.med.keio.ac.jp/
 and
 * GBF, Dept. of Genome Analysis,
 * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e-mail:
 info.genome@gbf.de
 * URL: http://genome.gbf.de/
 and
 * Max-Planck Institute for Molecular Genetics,
 * Innesstrasse 73, D-14195 Berlin, Germany,
 * e-mail: info-chr21@molgen.mpg.de
 * URL: http://chr21.rz-berlin.mpg.de/
 A1163240: Submitted (10-APR-2000).
 Location/Qualifiers
 1. 340000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 <1..46577
 /note="Accession No. AP001442"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q21.2"
 /clone="p1715, 5' partial"
 /clone_lib="CMP21 pl library"
 complement(1021..1140)
 /note="L1MA4"
 /rpt_family="LINE/L1"

```

repeat_region /rpt_type=DISPERSED
complement(1141. .1447)
/note="Aluuo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(1448. .1605)
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
1971. .2404
/note="MT1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(2487. .2779)
/note="Alusx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
3606. .3907
/note="Alusg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
4135. .4318
/note="MT1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
4319. .4607
/note="Alusg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
4608. .4835
/note="MT1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(5812. .6119)
/note="Aluuo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(6634. .6734)
/note="MER5A"
/rpt_family="DNA/MERL_type"
/rpt_type=DISPERSED
complement(8836. .9032)
/note="MER3A"
/rpt_family="DNA/MERL_type"
/rpt_type=DISPERSED
complement(9523. .9820)
/note="Alusg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(10813. .11086)
/note="L1MB3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(11095. .11195)
/note="U6"
/rpt_family="snRNA"
/rpt_type=DISPERSED
complement(11208. .11497)
/note="Alusg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(11498. .11780)
/note="Alusg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
13428. .13449
/note="(T)n"
/rpt_family="Simple_repeat"
/rpt_type=JANDEM
13483. .13746
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED

```

```

STS 13676. .13790
/standard_name="SHGC-31514"
/note="AFMa3562h9"
Accession NO. M3112"
/db_xref="GDB:9732389"
complement(13747. .14027)
/note="Aluub"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14028. .14430
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
14431. .14742
/note="Alusx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14743. .14806
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
15768. .16072
/note="Alusx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
17861. .18012
/note="Alr"
/rpt_family="SINE/Alr"
/rpt_type=DISPERSED
complement(18041. .18207)
/note="MER91A"
/rpt_family="DNA/MERL_type"
/rpt_type=DISPERSED
complement(18416. .18684)
/note="Alusg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(19782. .19899)
/note="L1MC5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
20230. .20272
/note="(CA)n"

```

```

Query Match 90.6%: Score 15.4; DB 90; Length 340000;
Best Local Similarity 76.5%: Pred. No. 7.8e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTYGARGARATGAYCC 17
DB 273372 TTTGAGGAATGATCC 273356

```

```

RESULT 21
G01707 355 bp DNA STS 07-DEC-1994
LOCUS CHICKEN STS ADL288.
DEFINITION G01707
ACCESSION G01707
VERSION 1 GI:595231
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE chicken vector-pbluescript II KS+ host-E. coll.
ORGANISM Gallus gallus
Eukaryote; Mitochondria; Metazoa; Chordata; Vertebrata; Tetrapoda;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 355)
AUTHORS Cheng, H. H.
JOURNAL Unpublished (1994)
SYNOPSIS: B407
CONTACT: Hans H. Cheng
AVIAN DISEASE AND ONCOLOGY LABORATORY
USDA-ARS
3606 E. Mount Hope Rd, East Lansing, MI 48823, USA
Tel: 517/3376758

```



```

misc_feature
/cclone="BD0AA004D05"
/cclone_id="BD0AA"
/notes="end : T7"
<77. >988
/notes="/similar to Saccharomyces cerevisiae ORF YHR05C [
GAL : GTP-binding protein alpha subunit of the pheromone
pathway 1"
/evidence="not experimental"

BASE COUNT      377 a      107 c      170 g      326 t      17 others

ORIGIN

Query Match      84.7%   Score 14.4;   DB 54;   Length 997;
Best Local Similarity 75.0%   Pred. No. 1.8e+03;
Matches 12;   Conservative 4;   Mismatches 0;   Indels 0;   Gaps 0;

QY      2 TYGARGARATGAYCC 17
Db      103 TTGAAGCAATGATGCC 88
1:||||:|||||:||||
1:||||:|||||:||||

RESULT 25
LOCUS      CNS070T9/c
DEFINITION CNS070T9      1084 bp      DNA      STS      11-JAN-2001
clone A20AA012B01 of library A20AA from strain CBS 712 of
Kluyveromyces marxianus, sequence tagged site.
ACCESSION      AL424051
VERSION      AL424051.1   GI:12207245
KEYWORDS      STS.
SOURCE      Kluyveromyces marxianus.
ORGANISM      Kluyveromyces marxianus
Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE      1 (bases 1 to 1084)
AUTHORS      Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.,
Dujon,B.
Genomic Exploration of the Hemiascomycetous Yeasts: 12.
Kluyveromyces marxianus var. marxianus
FEBS Lett. 487 (1), 71-75 (2000)
1152887
2 (bases 1 to 1084)
Souchet,T.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolojin-Pukhara,M., Bon,E., Brotier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durans,P., Leplongie,A., Llorente,B.,
Malpertuy,A., Neugeigise,C., Ozier-Kalogiropoulos,O., Potter,S.,
Saurin,M., Tekita,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
FEBS Lett. 487 (1), 3-12 (2000)
1152876
3 (bases 1 to 1084)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequenage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
candida, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1..1084
/organism="Kluyveromyces marxianus"
/strain="CBS 712"
/db_xref="taxon:4911"
/cclone="A20AA012B01"

```

BASE COUNT 262 a /clone_lib="AZ0AA"
ORIGIN 245 c 196 g 380 t 1 others

Query Match 84.7%; Score 14.4; DB 53; Length 1084;
Best Local Similarity 75.0%; Pred. No. 1.8e+03;
Matches 12; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGARATGGAGC 16
||:||||:||||:|
Db 426 TTCGAGAGAAATGAGTC 411

Search completed: May 15, 2001, 11:13:37
Job time: 3608 sec